

Keeping the Maltese Language Alive: AI Chatbot Breaks Through Bilingual Boundaries





Author: **Christian Keszthelyi**

Researchers at UM are developing a chatbot that communicates in Maltese and can carry out a conversation flawlessly as if it were human. The project is unique, as previously most chatbots handling Maltese were rule-based, meaning their interactions were limited to predefined menus.

Working with the Maltese language is challenging. Unlike

English, building a Maltese digital corpus (a large, structured set of texts) is difficult due to the restricted availability of resources – both textual and human labour.

Dr Claudia Borg, from UM's Department of Artificial Intelligence, is working on this titanic task through the framework of The New Era of Chatbot Project. Borg leads the project with Dr Marc Tanti (Institute of Linguistics and Language Technologies) and Prof. Michael Spagnol (Department of Maltese). Two Ph.D. students, Kurt Micallef and Kurt Abela, and a postdoctoral student, Dr Marthese Borg, are also contributing to this task. The team's main aim is to create a conversational Maltese chatbot, so that instead

of predetermined menus and rigid prompts, users can converse with the software in a more natural form.


WHAT IT TAKES

Training a chatbot to move beyond predefined prompts happens through immersion in real-life scenarios. 'We developed user stories related to banking and finance, envisioning scenarios where users might inquire about opening a bank account in Maltese. We trained the chatbot to understand various ways users might phrase their questions and how to respond,' says Borg.

The chatbot prototype has already been tested in controlled lab environments. In this step, the software is exposed to preset, real-life scenarios so it can learn how requests should be handled and start becoming conversational. The next step is interacting in a real-world setting to allow broader public

usage. However, such a trial comes with challenges, as people express themselves in diverse ways in writing, and grammatical or typographical errors can occur. Still, the researchers are confident that their preparation and extensive research in Maltese language processing will help the chatbot perform well.

'Ultimately, our mission is to ensure that Maltese remains relevant in today's technology-driven communication landscape. We commit ourselves to developing the necessary technology to keep the Maltese language vibrant in the digital sphere, alongside other ongoing efforts,' Borg says.

The training happens in an exciting language environment, as Maltese natives are practically bilingual and often switch between Maltese and English in the same sentence. Users might say, 'Kif nistgħu nifthu kont?' the Maltese equivalent of 'How 



Dr Claudia Borg
Photo by James Moffett



Prof. Michael Spagnol
Photo by James Moffett



Dr Marc Tanti
Photo by James Moffett

can we open an account?' But, in a country that has been bilingual for decades, someone may use both languages in their query, saying: 'Kif nistgħu nifthu account?' This would involve elements of code-switching, the practice of alternating between two or more languages or language varieties in a conversation or within a single utterance. For an older, rule-based chatbot, this may have caused issues. Not so for the chatbot in Borg's project, though.

'We are training the chatbot to recognise these different expressions and potential spelling mistakes. While we aim to be creative in our training, we understand it won't be foolproof. After the launch, we expect unforeseen challenges, such as unanticipated spelling errors,' Borg tells **THINK**.

The project aims to ensure that the technology they develop understands how people mix and express themselves in Maltese (and English). Nevertheless, the chatbot prioritises responding in proper Maltese, free of grammatical or spelling errors, which is crucial for effective communication.

LEARNING TO SPEAK

Language models have been evolving in size, particularly in

Maltese. Currently, the researchers are working with a BERT-based language model. Bidirectional Encoder Representations from Transformers, or BERT, is a model developed by Google that has been influential in natural language processing tasks. But BERT predates technologies like ChatGPT, whose models are large neural networks that rely on vast amounts of data. ChatGPT's neural network of English far exceeds the digital corpus that is currently available in the Maltese language.

'Our ongoing experiments aim to develop a Maltese large language model (LLM), using open-source models that we further train with Maltese data to enhance integration. Although we are in the early stages and have seen some promising results, creating an LLM for Maltese is expensive,' Borg says.

Researchers need to balance research goals with environmental impact when training these models. Artificial intelligence – especially as the public has recently come to know it: generating text, images, and videos in seconds – is a power-intensive technology. Every processing task is fuelled by immense computational power using graphics processing units (GPUs), which

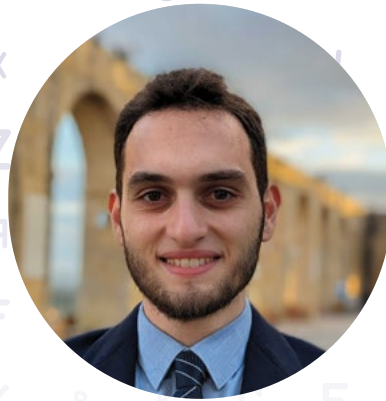
consume natural resources through electricity. Researchers' computing resources are finite, and they have to make sure that whenever they use computing power, it is used wisely. Many LLMs require significant GPU power to operate. For Borg, it is imperative that the Maltese model they are working on is not just a theoretical exercise but one that has practical applications and cost-effectiveness in mind.

'Given our data and computing power limitations, we focus on maximising existing resources. We prefer smaller, multilingual models that can leverage cross-lingual transfer, where a model can be trained on similar languages like Italian or Arabic. This strategy allows us to use less data in Maltese while achieving effective results by incorporating knowledge from similar languages,' Borg says.

But shouldn't ChatGPT (by OpenAI) be helpful for such research, since you can speak to it in Maltese and it will answer? Well, one of the main issues with ChatGPT is that it is closed-source, and OpenAI retains ownership of the code. 'We generally prefer open-source models versus closed-source. Moreover, small industries in Malta need economically



Kurt Abela
Ph.D. student



Kurt Micallef
Ph.D. student



Dr Marthese Borg
Postdoctoral student

viable solutions to integrate Maltese-language technologies,' says Borg.

A significant challenge with Maltese is the lack of established benchmarks to evaluate model performance. Unlike English, which has numerous benchmarks across various tasks, Maltese benchmarks are limited and primarily available for more straightforward tasks like machine translation. The researchers are currently working on creating datasets to test and assess models on more complex tasks, such as summarisation, to improve language model performance for Maltese.

BATTLING FOR MALTESE

The chatbot aims to serve an essential, nobler purpose than one might assume. In Malta, the English language has lived in symbiosis with Maltese, and for this, the native tongue has suffered. Maltese might very well lose relevance within our society against the utility of speaking English. The chatbot the UM team is working on could help Maltese speakers use their native language in more scenarios, thus keeping it alive.

'Maltese should be seen as a language for reading and writing, not just casual conversation. Making technologies available in Maltese

helps reinforce its use, though it's not the sole solution to keeping the language alive. The goal should be to make using Maltese the default choice – something that happens naturally, without needing encouragement,' Borg says.

But as English has been adopted as the native language of international communication, how much hope is there that Maltese will not suffer further in this tiny nation of roughly half a million native speakers?

'I'm optimistic about the future of the Maltese language, and history gives me good reason to be. When Malta joined the European Union, Maltese became an official EU language, which led to a surge in literature, novels, and translations in Maltese. Suddenly, we saw more works by both professional translators and Maltese authors writing original content. This change created more opportunities for using Maltese, even though the immediate impact was on a relatively small group – authors, translators, and publishers. However, it showed how making Maltese more visible could inspire people to use it more confidently,' Borg says.

Intuitive, well-working digital tools available in Maltese can have a similar

impact. The more integral the Maltese language becomes to its speakers in everyday digital experiences, the more it becomes a default choice for communication. It is an unconscious shift, encouraging people to think in Maltese. Native Maltese speakers would then feel more comfortable not just speaking in Maltese but also writing in Maltese, whether for emails or other forms of communication.

'The real difference between Maltese and English isn't the languages themselves but the digital tools and resources available to support them. English dominates because of its accessibility through TV programs, apps, and platforms. By making Maltese just as accessible and visible in the digital world, we can create an environment where using Maltese is effortless and intuitive,' Borg concludes. **T**

The New Era of Chatbot Project is financed by a Research and Development Grant funded by Malta Enterprise. Cartesio Ltd is the lead project manager, and Noovle International, a subsidiary of TIM Enterprise, is the technical implementer. The University of Malta is responsible for the research component of the project.